

ICTNET at Session Track TREC 2012

Zhenhong Chen^{1,2}, Mingchuan Wei^{1,2}, Junxiao Nan^{1,2}, Jun Chen^{1,2}

Xiaoming Yu¹, Yue Liu¹, Xueqi Cheng¹

1. Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100190

2. Graduate School of Chinese Academy of Sciences, Beijing, 100190

{chenzhenhong, weimingchuan, nanjunxiao, chenjun, yuxiaoming}@software.ict.ac.cn

{liuyue, cxq}@ict.ac.cn

1 Introduction

In this paper, we describe our solutions to the Session Track at TREC 2012. The main contribution of our work is that we implement the learning to rank model to re-rank the documents retrieved by our search engine[2]. We notice that Huurnink et al. [3] have used learning to rank algorithm to model session features at last year's Session Track. Due to lacking of training data, their model did not outperform substantially than others. Intuitively, we use last year's session data for tuning the weights of ranking features. Meanwhile, we define several useful features to model session search intent.

The rest of this paper is organized as follows. We detail our models in section 2. Section 3 describes our experiments, including retrieve system setup, our research structure and our evaluation results. Conclusions are made in the last section.

2 Our approach

In our work, we pose several methods utilizing the session information to improve search engine results. It should be noted that all our methods are used to re-rank the documents retrieved by our search engine only using each session's last query, unlike most of other participants, who use variant query expansions to get results from Indri. Details of our search engine setup are described in section 3.1.

2.1 Query Expansion

We use the previous queries in the same session to extend the current query. The final query consists of all the terms both in the historical queries and the current query. Let q_1 to q_{m-1} stand for previous queries and q_m stands for the current query, then $p(w|q)$ denotes the weight of the word w in final query q , which is calculated as (1).

$$p(w|q) = \exp(d * \max_i (w \in q_i)) \quad (1)$$

Aiming to enhance the importance of the terms occur in the latter query, we set the control parameter d as 0.05.

We re-rank the documents by calculating the Weighted-BM25 score between the extended query and the retrieved documents.

2.2 Virtual Document Model

Obviously, the ranked list returned by the search engine can serve as a good profile of the search intent. We can use this contextual information to construct the so-called virtual document to model the information need of the user. We simply incorporate the titles and snippets of each past query together. Then we use the cosine similarity between the retrieved document and the virtual document to re-rank the results. Depending on the source of the ranked list, we develop two methods. Firstly, we use the given session data from the NIST to construct the virtual document. Secondly, we submit the current query of each session to Google and obtain the virtual document based on the first 24 results.

Report Documentation Page			Form Approved OMB No. 0704-0188		
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE NOV 2012		2. REPORT TYPE		3. DATES COVERED 00-00-2012 to 00-00-2012	
4. TITLE AND SUBTITLE ICTNET at Session Track TREC 2012			5a. CONTRACT NUMBER		
			5b. GRANT NUMBER		
			5c. PROGRAM ELEMENT NUMBER		
6. AUTHOR(S)			5d. PROJECT NUMBER		
			5e. TASK NUMBER		
			5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Chinese Academy of Sciences, Institute of Computing Technology, Beijing, 100190,			8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)			10. SPONSOR/MONITOR'S ACRONYM(S)		
			11. SPONSOR/MONITOR'S REPORT NUMBER(S)		
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES Presented at the Twenty-First Text REtrieval Conference (TREC 2012) held in Gaithersburg, Maryland, November 6-9, 2012. The conference was co-sponsored by the National Institute of Standards and Technology (NIST) the Defense Advanced Research Projects Agency (DARPA) and the Advanced Research and Development Activity (ARDA). U.S. Government or Federal Rights License					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as Report (SAR)	18. NUMBER OF PAGES 4	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

2.3 Optimization Based on User's Attention Time

In last year's Session Track, BUPT_WILDCAT team achieved the best results in RL4. We simply implement their model [4] as one of our runs. Parameters are the same value as the BUPT_WILDCAT team used in their work.

2.4 Learning to Rank

In this section, we detail the main contribution of our work, using machine learning to model the user's search intent. We apply the SVM^{rank} [5] algorithm to learning from explicit relevance judgments on last year's Session Track data. The features we used in our submitted runs are listed in table 1.

Table 1: features used in SVM^{rank}

Feature	feature description
PageRank	the PageRank of the document, derive from [6]
QE	the score of Query Expansion Model
SessionVD	the score of Session Virtual Document Model
CAT	the score of Optimization Based on User's Attention Time Model
CosSimQT	cosine similarity between query and title
BM25QC	BM25 score between query and content
GoogleVD	the score of Google Virtual Document Model

We use various combinations of the features to generate our learning to rank model for different runs. Since Google Virtual Document Model may be consider as using external resource, we design two separate runs, one of which uses the GoogleVD feature, while the other not uses it.

3 Experiments

We have conducted experiments to verify the effectiveness of our models. In this section, we first describe the search engine and search strategy we used to retrieve the result set of each session. Then we detail all our submissions and evaluation results.

3.1 Experiment Setup

We submit the last query of each session to Golaxy[2]. Our search strategy is to retrieve the initial document set that satisfy the condition that the content filed of each document contains all the terms of the query or the title field contains at least one term, without any stop word removal or stemming. Then we use Waterloo spam ranking score [7] to filter documents with "fusion" spam score [8] less than 70%. Finally, we apply the BM25 model to rank the previous results.

3.2 Our Runs

We have submitted three runs in this year's Session Track. All our submitted runs are Category A runs. The research structure and models implemented in each runs are listed in Table 2.

Table 2: Methods in all runs

Run ID	ICTNET12SER1	ICTNET12SER2	ICTNET12SER3
RL1	Result of Golaxy	Google Virtual Document	SVM^{rank} , features: PageRank, CosSimQT, BM25QC, GoogleVD
RL2	Query Expansion	SVM^{rank} , features: PageRank, CosSimQT, BM25QC, QE	SVM^{rank} , features: PageRank, CosSimQT, BM25QC, QE, GoogleVD
RL3	Session Virtual Document	SVM^{rank} , features: PageRank, CosSimQT, BM25QC, QE,	SVM^{rank} , features: PageRank, CosSimQT,

		SessionVD	BM25QC, QE, SessionVD, GoogleVD
RL4	Optimization Based on User's Attention Time	SVM ^{rank} , features: PageRank, CosSimQT, BM25QC, QE, SessionVD, CAT	SVM ^{rank} , features: PageRank, CosSimQT, BM25QC, QE, SessionVD, CAT, GoogleVD

3.3 Evaluation Results

Evaluation results of our submissions at 2012 Session Track are showed in Table 3. The highest score for each experimental condition is indicated in bold. According to Table 3, we can conclude that our models can significantly improve the performance of a search engine when take advantage of session information. We obtain 80.14% of performance increase when compare our best result (ICTNET12SER3.RL4) with the direct result from our search engine (ICTNET12SER1.RL1). Again, we want to emphasize that all our methods are used to re-rank the documents retrieved by our search engine. In Figure 1, we simply compare ICTNET12SER3.RL4 with the median result of RL4 of all participators and observe that nearly 80% of the sessions outperform the median result.

Table 3: Results on 2012 Session Track, in terms of NDCG@10

	RL1	RL2	RL3	RL4
ICTNET12SER1	0.1586	0.2043	0.2039	0.2392
ICTNET12SER2	0.2144	0.2168	0.2732	0.2827
ICTNET12SER3	0.2481	0.2476	0.2640	0.2857

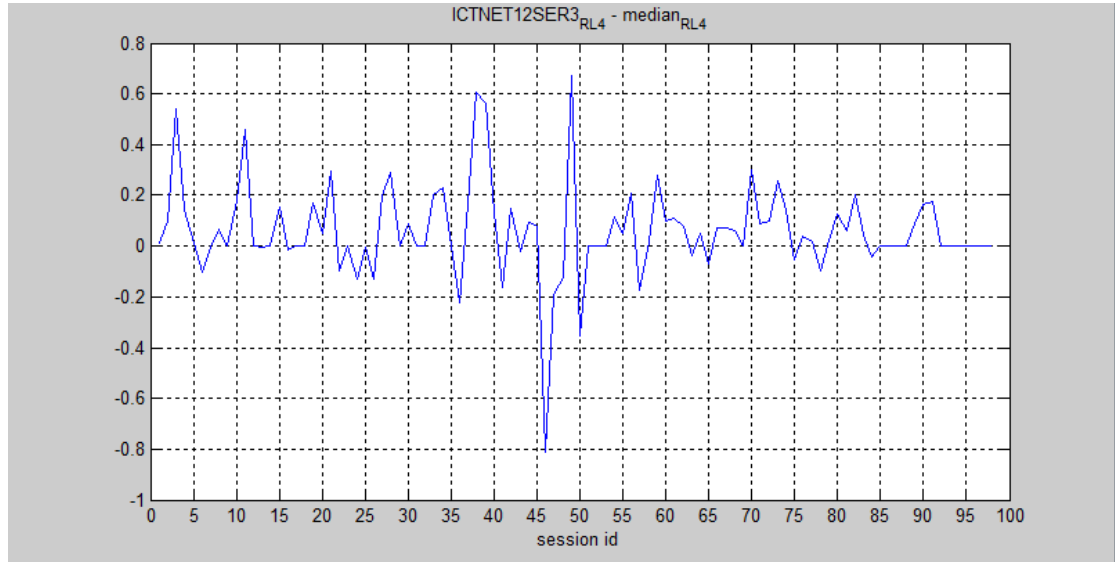


Figure 1: Comparison between ICTNET12SER3_RL4 and Median_RL4 on ndcg@10 for all sessions, nearly 80% of the sessions outperform the median result

4 Conclusions

In this paper, we presented several approaches to verify whether a retrieve system can use increasing amounts of information prior to a query to improve effectiveness for that query. Each one of our methods models a special aspect of the relationship between the query and the corresponding document. Thus, when combining all these models with a learning to rank algorithm, we can expect significantly improvement in effectiveness. Experiment results confirm

our expectation impressively. We have achieved 80.14% of performance increase compared with the result from our search engine.

For the future work, we will consider incorporating more features to model user's search intent, including URL and anchor information. Besides, we will investigate other learning to rank algorithms and similarity measures. Feature selection and parameter optimization will also be applied to achieve better performance.

5 Acknowledgements

We would like to thank all organizers and assessors of TREC and NIST. This work is sponsored by NSF of China Grants No. 60933005, No. 61100083 and No.61173064, and by 242 Program of China Grants No.2011F65.

6 References

- [1] Session Track 2011 Overview
- [2] ICTNET at Web Track 2011 Ad-Hoc Track
- [3] The University of Amsterdam at the TREC 2011 Session Track
- [4] BUPT_WILDCAT at TREC 2011 Session Track
- [5] T. Joachims, Training Linear SVMs in Linear Time, Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD), 2006.
- [6] <http://boston.lti.cs.cmu.edu/clueweb09/wiki/tiki-index.php?page=PageRank>
- [7] Gordon V. Cormack, Mark D. Smucker, and Charles L. A. Clarke. Efficient and effective spam filtering and re-ranking for large web datasets. Information Retrieval, 14(5), 2011: 441-465.
- [8] <http://plg.uwaterloo.ca/~gvcormac/clueweb09spam/>